

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:36 ; Search time 30.11 Seconds  
(without alignments)  
464.806 Million cell updates/sec

Title: US-09-863-063-2

Perfect score: 682

Sequence: 1 AQSVPGGDIQTQPGTKIVFN.....EWFQDGMVRRKNLPIEYNP 126

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*

1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	114	16.7	242	21 AAY94850	Human protein clon
2	113	16.6	236	20 AAW86813	Human vesicular bi
3	113	16.6	300	20 AAY76635	Human ovarian tumo
4	105.5	15.5	341	22 AAO12925	Human polypeptide
5	101	14.8	626	22 ABG01123	Novel human diagn
6	99	14.5	308	22 AAU30212	Novel human secret
7	98	14.4	76	22 AAU20373	Human secreted pro
8	95	13.9	223	22 AAU01095	Gene 27 Human secr
9	95	13.9	243	20 AAY41727	Human PRO983 prote
10	95	13.9	243	21 AAB44283	Human PRO983 (UNQ4
11	95	13.9	243	22 AAU29079	Human PRO polypept



21-OCT-1999.  
09-APR-1998; 98DE-1017557.  
09-APR-1998; 98DE-1017557.  
(META-) METAGEN GES GENOMFORSCHUNG MBH.  
Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
WPI; 1999-591920/51.  
N-PSDB; AAZ77504.  
New nucleic acid sequences expressed in ovarian, and some other, cancer  
tissues, and derived polypeptides, for treatment of ovarian cancer and  
identification of therapeutic agents -  
Claim 25; Page 298; 310pp; German.  
This invention describes novel nucleic acid (cDNA) sequences (A) which  
have anticancer activity and are highly expressed in ovarian tumor  
tissue (and some also in testis and breast cancer tissue). The products  
of the invention can be used for gene therapy. (A) are used (i) for  
recombinant expression of polypeptides (B) and (ii) to isolate complete  
genes. (B) are used (i) to identify agents suitable for treatment of  
ovarian cancer; (ii) directly for treating this form of cancer  
(including expression from gene therapy vectors) and (iii) for generation  
of specific antibodies. (A) are identified by assembling ESTs (expressed  
sequence tags) from a particular tissue type before comparison of  
expression patterns. This allows a significantly longer fragment of the  
gene to be revealed, so should reduce the number of failures associated  
with the fact that ESTs from different libraries may represent different  
parts of the same unknown gene, distorting the estimated frequency of  
occurrence in a particular tissue. AAY76505-Y76638 represent protein  
fragments encoded by the human ovarian tumor cDNA library derived EST  
fragments represented in AAZ77450-Z77572.  
Sequence 300 AA;  
Query Match 16.6%; Score 113; DB 20; Length 300;  
Best Local Similarity 28.2%; Pred. NO. 2.3e-05;  
Matches 24; Conservative 19; Mismatches 42; Indels 0; Gaps 0  
QY 13 PGTKIVFNAPYDDKHTYHIKVINSGARRIGYGIKTTNMRILGVDPGVLDPKEAVLLAV 72  
Db 70 pptdlkfgpftdvvttnlkrnpdrkvckvktaprrycvrpnsgildpgstvtvs 129  
QY 73 SCDAFAFGQEDTNNDRITVEWTNTP 97  
Db 130 mlqffdydpnekskhkfmvqtifap 154  
RESULT 4  
ID AAO12925  
XX AAO12925 standard; Protein; 341 AA.  
AC AAO12925;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 26817.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.





XX PF 17-JAN-2001; 2001WO-US01347.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-451931/48.  
XX DR N-PSDB; AAS33082.  
XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing  
PT or treating medical conditions -  
XX PS Claim 11; SEQ ID No 365; 753pp; English.  
XX CC The invention relates to novel isolated nucleic acid molecules (I)  
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,  
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,  
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in  
CC the prevention, treatment and diagnosis of diseases associated with  
CC inappropriate expression of secreted proteins. (I) and complementary  
CC sequences may also be used as DNA probes in diagnostic assays (e.g.  
CC of similar nucleic acid sequences (PCR)) to detect and quantitate the presence  
CC be in need of restorative therapy. (II) may also be used as antigens in  
CC the production of antibodies and in assays to identify modulators  
CC (agonists and antagonists) of the expression and activity of the secreted  
CC proteins. The anti-(II) antibodies and antagonists may also be used to  
CC down regulate expression and activity of (II). The anti-(II) antibodies  
CC may also be used as diagnostic agents for detecting the presence of (II)  
CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The  
CC disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/  
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,  
CC angina and thrombosis), infections caused by bacteria, viruses and  
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),  
CC agonists, antagonists and antibodies can also be used to promote wound  
CC healing, maintain organs before transplantation, and support cell culture  
CC of primary tissues. AAU20342-AAU20666 represent human secreted protein  
CC amino acid sequences, and related sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO  
CC at: ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 76 AA;

Query Match 14.4%; Score 98; DB 22; Length 76;  
Best Local Similarity 32.8%; Pred. No. 0.00025;  
Matches 22; Conservative 14; Mismatches 29; Indels 2; Gaps 1;

QY 15 TKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKAEVLL--AV 72  
Db 9 tdxkfgpftdvvttnlklrnpdrkvcfkvktaprrycvrrpnsqildpgstvtvsgsk 68  
QY 73 SCDAFAF 79  
Db 69 scsefmy 75

RESULT 8  
AAU01095  
ID AAU01095 standard; Protein; 223 AA.  
XX AC AAU01095;  
XX DT 18-JUL-2001 (first entry)

XX DE Gene 27 Human secreted protein homologous amino acid sequence.  
XX KW Human secreted protein; gene therapy; autoimmune disease;  
KW hyperproliferative disorder; cardiovascular disorder;  
KW cerebrovascular disorder; nervous system disorder; infection;  
KW ocular disorder; wound healing; epithelial cell proliferation;  
KW skin aging; transplantation; tissue regeneration; chemotaxis;  
KW food additive; preservative.  
XX OS Homo sapiens.  
XX PN WO200123402-A1.  
XX PD 05-APR-2001.  
XX PF 26-SEP-2000; 2000WO-US26376.  
XX PR 27-SEP-1999; 99US-0155808.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI; 2001-266138/27.  
XX PT Nucleic acids encoding 43 human secreted polypeptides, useful for  
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
PT disease and diabetic retinopathy -  
XX PS Disclosure; Page 499-500; 516pp; English.  
XX CC AAU01083-AAU01105 represent human secreted protein homologous amino acid  
CC sequences. They are described in the invention of 43 novel human secreted  
CC proteins (AAU01040-AAU01082) and their gene sequences (AAS01460-AAS01502)  
CC which can be used in gene therapy. The secreted proteins are useful to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The secreted  
CC proteins are also useful in diagnosing a pathological condition or  
CC susceptibility to a pathological condition. Antibodies to the secreted  
CC proteins can also be used in alleviating symptoms associated with  
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or  
CC enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed  
CC or treated include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Parkinson's disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities.

XX SQ Sequence 223 AA;

Query Match 13.9%; Score 95; DB 22; Length 223;  
Best Local Similarity 23.4%; Pred. No. 0.0026;  
Matches 22; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

QY 9 IQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKAEV 68  
Db 8 lslepqlhelkfrgpfddvvttnlklrnpdrkvcfkvktaprrycvrrpnsqildagasi 67  
QY 69 LLAVSCDAFAFAGQEDTNNDRITVEW---TNTPD 98  
Db 68 nvsmlqpfdydpnekskhkfmvqsmfapttdtsd 101

RESULT 9  
AAU41727







PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.

XX (GETH ) GENENTECH INC.

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

XX WPI; 2001-602746/68.  
DR N-PSDB; AAS45980.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and  
PT to screen for modulators of the compounds -

XX Claim 11; Fig 112; 774pp; English.

XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
CC The PRO polypeptides and their associated nucleic acids can be used to  
CC detect the presence of a tumour in a mammal by comparing the level of  
CC expression of a PRO polypeptide in a test sample of cells from the animal  
CC and a control sample of normal cells, whereby a higher level of  
CC expression in the test sample indicates the presence of a tumour in the  
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
CC and rabbits but are preferably human. The polypeptides can be used to  
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
CC when contacted with it. A specific polypeptide can be used to stimulate  
CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 243 AA;

Query Match 13.9%; Score 95; DB 22; Length 243;

Best Local Similarity 23.4%; Pred. No. 0.003;

Matches 22; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

QY 9 IQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAV 68

Db 8 lslepqhelkfrgpfddvvttnlklgnptdrnvcfkvktaprrycvrpnsgiidagasi 67

QY 69 LLAVSCDAFAFGQEDTNNDRITVEW---TNTPD 98

Db 68 nvsvmlqpfidydpnekskhkfmvqsmfapttdtsd 101

RESULT 12

AAAB953358

ID AAB953358 standard; Protein; 243 AA.

XX AC AAB953358;

XX AC AAB953358;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17658.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

XX Claim 8; SEQ ID 17658; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 243 AA;

Query Match 13.9%; Score 95; DB 22; Length 243;

Best Local Similarity 23.4%; Pred. No. 0.003;

Matches 22; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

QY 9 IQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAV 68

Db 8 lslepqhelkfrgpfddvvttnlklgnptdrnvcfkvktaprrycvrpnsgiidagasi 67

QY 69 LLAVSCDAFAFGQEDTNNDRITVEW---TNTPD 98

Db 68 nvsvmlqpfidydpnekskhkfmvqsmfapttdtsd 101

RESULT 13

AAAW31324

ID AAW31324 standard; peptide; 16 AA.

XX AC AAW31324;

XX 13-MAR-1998 (first entry)

XX Dictyocaulus viviparus DV18 peptide fragment 4.  
XX Immunogenic protein; DV18; lungworm; vaccine; immunity;  
XX dictyocauliasis; cattle; immunoassay.  
XX Dictyocaulus viviparus.



XX EP785253-A1. 99US-0130510.  
PN 23-APR-1999; 99US-0130891.  
XX 23-JUL-1997. 99US-0131449.  
PD 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
PF 27-DEC-1996; 96EP-0120947. 99US-0132484.  
XX 19-JAN-1996; 96DE-4001754. 99US-0132485.  
XX (FARH ) HOECHST AG. 99US-0132486.  
PA Hofmann J, Schmid K; 99US-0132487.  
PI WPI; 1997-365928/34. 99US-0132486.  
XX Immunogenic Dictyocaulus viviparus lungworm protein - for use in  
DR vaccines and immunoassays 99US-0132486.  
PT Example 10; Page 6; 17pp; German. 99US-0132487.  
XX This sequence represents a peptide fragment of the immunogenic protein  
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid  
CC can be joined to the amino acids Asp, Asn, Ser, Gly or Arg. DV18 can be  
CC used in vaccines for immunising cattle against dictyocauliasis and in an  
CC ELISA immunoassay for determining DV18-specific antibodies in the blood  
CC of cattle. 99US-0132487.  
XX Sequence 16 AA; 99US-0132487.  
SQ Query Match 12.5%; Score 85; DB 18; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 104 FRREWFQDGMVRRK 118  
Db 2 frrewfgdgmvrk 16  
RESULT 14  
AAG48788  
ID AAG48788 standard; Protein; 204 AA.  
XX  
AC AAG48788;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61647.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
XX 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PT 19-MAY-1999; 99US-0135124.  
XX 20-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
XX 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
CC 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137222.  
CC 01-JUN-1999; 99US-0137528.  
CC 03-JUN-1999; 99US-0137502.  
CC 04-JUN-1999; 99US-0137724.  
XX 07-JUN-1999; 99US-0138094.  
SQ 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
XX 10-JUN-1999; 99US-0139119.  
PR 14-JUN-1999; 99US-0139453.  
XX 16-JUN-1999; 99US-0139492.  
PR 17-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
XX 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
XX 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
XX 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
XX 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
XX 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
XX 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
XX 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
XX 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
XX 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
XX 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
XX 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
XX 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
XX 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
XX 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.

Search completed: September 20, 2002, 09:23:16  
Job time: 40 sec.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 25.65 seconds  
(without alignments)  
119.985 Million cell updates/sec

Title: US-09-863-063-2  
Perfect score: 682  
Sequence: 1 AQSVPVPGDIQTQPGTKIVFN.....EWFQGDGMVRRKNLP1EYNP 126

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	593	87.0	127	4	US-08-786-455B-14
2	134	19.6	260	3	US-08-857-213-3
3	113	16.6	236	3	US-08-857-213-1
4	76	11.1	20	4	US-08-786-455B-2
5	66	9.7	1003	1	US-08-571-758-4
6	66	9.7	1003	1	US-08-909-984A-4
7	66	9.7	1003	1	US-08-909-983-4
8	65.5	9.6	890	2	US-08-483-101-14
9	64.5	9.5	14	4	US-08-786-455B-4
10	63.5	9.3	305	2	US-08-853-659A-47
11	62.5	9.2	139	2	US-08-553-501A-29
12	62.5	9.2	139	3	US-09-205-231-29
13	62	9.1	1076	2	US-08-867-941-19
14	62	9.1	1076	4	US-09-074-658-19
15	61.5	9.0	405	5	PCT-US93-11404-2
16	61	8.9	995	2	US-08-673-789-5
17	61	8.9	1013	2	US-08-866-650-3
18	61	8.9	1013	2	US-09-021-287-3
19	61	8.9	1013	4	US-09-240-473-3
20	61	8.9	1864	2	US-08-804-227C-3
21	60.5	8.9	120	2	US-08-553-501A-88
22	60.5	8.9	120	3	US-09-205-231-88
23	60.5	8.9	312	4	US-09-475-316A-48
24	60.5	8.9	312	4	US-09-475-316A-50
25	60.5	8.9	312	4	US-09-475-316A-52
26	60.5	8.9	312	4	US-09-475-316A-54
27	60.5	8.9	312	4	US-09-475-316A-56

28	60.5	8.9	312	4	US-09-475-316A-58
29	60.5	8.9	979	3	US-08-870-529-2
30	60	8.8	113	2	US-08-256-568B-83
31	60	8.8	113	4	US-09-038-369B-83
32	59.5	8.7	246	1	US-08-197-834-7
33	59.5	8.7	735	2	US-08-313-185-48
34	59.5	8.7	735	2	US-08-459-499-9
35	59.5	8.7	735	2	US-08-459-499-12
36	59.5	8.7	735	3	US-09-082-614A-48
37	59.5	8.7	1288	3	US-08-762-428A-6
38	58	8.5	897	1	US-08-095-737-4
39	58	8.5	897	1	US-08-480-145-4
40	58	8.5	897	2	US-08-477-389-4
41	57.5	8.4	172	2	US-08-853-659A-46
42	57.5	8.4	249	4	US-09-345-468-9
43	57.5	8.4	319	4	US-09-345-468-5
44	57.5	8.4	339	4	US-09-345-468-3
45	57.5	8.4	654	1	US-08-392-828C-2

ALIGNMENTS

RESULT 1

US-08-786-455B-14  
; Sequence 14, Application US/08786455B  
; Patent No. 6193971  
; GENERAL INFORMATION:  
; APPLICANT: HOFMANN, Joachim  
; APPLICANT: SCHMID, Karlheirich  
; TITLE OF INVENTION: DICTYOCAULUS VIVIPARUS ANTIGEN FOR  
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,455B  
; FILING DATE: 21-JAN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 196 01 754.8  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GRANADOS, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 18748/327  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 127 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-786-455B-14

Query Match 87.0%; Score 593; DB 4; Length 127;  
Best Local Similarity 86.2%; Pred. NO. 1.1e-67;  
Matches 106; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVL 62  
 Db 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVL 62  
 QY 63 DPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKNLPI 122  
 Db 63 DPKEATLMAVSCDTFEYGREDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKNLPI 122  
 QY 123 EYN 125  
 Db 123 EYN 125

RESULT 2  
 US-08-857-213-3  
 ; Sequence 3, Application US/08857213  
 ; Patent No. 6054290  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; APPLICANT: Hawkins, Phillip R.  
 ; APPLICANT: Murry, Lynn E.  
 ; TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/857,213  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0297 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 260 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: 1000369

Query Match 19.6%; Score 134; DB 3; Length 260;  
 Best Local Similarity 29.9%; Pred. No. 6e-09;  
 Matches 29; Conservative 18; Mismatches 50; Indels 0; Gaps 0;  
 QY 12 QPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKEAVLLA 71  
 Db 11 EPAGELRFKGFPTDVVTTADLKLNSPTDRRCFKVKTAPKRYCVRPNPNSGILEPKTSIAVA 70  
 QY 72 VSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREW 108  
 Db 71 VMLQPFNYDPNEKNKHKFNVQSMYAPDHVVESQELLW 107

RESULT 3  
 US-08-857-213-1  
 ; Sequence 1, Application US/08857213  
 ; Patent No. 6054290  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; APPLICANT: Hawkins, Phillip R.  
 ; APPLICANT: Murry, Lynn E.  
 ; TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/857,213  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0297 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 236 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: FIBRNGT01  
 ; CLONE: 148415  
 ; US-08-857-213-1

Query Match 16.6%; Score 113; DB 3; Length 236;  
 Best Local Similarity 28.8%; Pred. No. 2.4e-06;  
 Matches 23; Conservative 19; Mismatches 38; Indels 0; Gaps 0;  
 QY 13 PGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKEAVLLAV 72  
 Db 12 PPTDLKFKGPFPTDVVTTNLKLRNPSDRKVCVKVKTAPKRYCVRPNNSGIIIDPGSTVTVSV 71  
 QY 73 SCDAFAFGQEDTNNDRITVE 92  
 Db 72 MLQPFYDNDNEKSKHKFMVQ 91

RESULT 4  
 US-08-786-455B-2  
 ; Sequence 2, Application US/08786455B  
 ; Patent No. 6193971  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HOFMANN, Joachim  
 ; APPLICANT: SCHMID, Karlheirich  
 ; TITLE OF INVENTION: DICTYOCALUS VIVIPARUS ANTIGEN FOR  
 ; DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION



```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-909-984A-4

Query Match          9.7%; Score 66; DB 1; Length 1003;
Best Local Similarity 21.0%; Pred. No. 18;
Matches 26; Conservative 19; Mismatches 45; Indels 34; Gaps 4;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSA-----RRIGYGIKTTNM 50
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 602 SLTGSQVSTHSATSQVSNVSGSSSATYTSSLVNSGFFPRKLSNAGVDKRVPTSEYTD 661
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 51 KRLGVDPGVLDPKPEAVLLAVSCDAFAFGQEDTNNDRTITVEWTNTPDGAARQFR----- 105
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 662 HKSN-----DSDKTVSLS-----GSASTDSDRTPVRLDSTEDGDSGQWRQNSIS 705
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 -REW 108
   : ||
Db 706 LKEW 709

RESULT 7
US-08-909-983-4
; Sequence 4, Application US/08909983
; Patent No. 5747288
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wassarman, David A.
; TITLE OF INVENTION: A No. 5747288el Protein Kinase Required for Ras
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,983
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,758
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-909-983-4
```

```
Query Match          9.7%; Score 66; DB 1; Length 1003;
Best Local Similarity 21.0%; Pred. No. 18;
Matches 26; Conservative 19; Mismatches 45; Indels 34; Gaps 4;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSA-----RRIGYGIKTTNM 50
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 602 SLTGSQVSTHSATSQVSNVSGSSSATYTSSLVNSGFFPRKLSNAGVDKRVPTSEYTD 661
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 51 KRLGVDPGVLDPKPEAVLLAVSCDAFAFGQEDTNNDRTITVEWTNTPDGAARQFR----- 105
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 662 HKSN-----DSDKTVSLS-----GSASTDSDRTPVRLDSTEDGDSGQWRQNSIS 705
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 -REW 108
   : ||
Db 706 LKEW 709

RESULT 8
US-08-483-101-14
; Sequence 14, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-483-101-14
```

```
Query Match          9.6%; Score 65.5; DB 2; Length 890;
Best Local Similarity 28.8%; Pred. No. 17;
Matches 32; Conservative 14; Mismatches 40; Indels 25; Gaps 6;

QY 8 DIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG--VLDPK 65
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 DVFVQAGDIINDNGRYVEKQKNHRSAINS-----GLRLPLTRNLAVQLCCGGAVIDNK 425
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 -----EAVLLAVSCDA---FAFGQEDTNNDRTITVEWTNTPDGAARQF 104
   : || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```





QY 8 DIQTQPGTKIVFNAPYDDKHT-----YHIKV-----INSSARRIGYGI 45

Db 474 DVNKPFSVKEVDNAYKEQHNLIKAVFNKMKALGSTHHHINLQVGYDKFNSSLSRVEYRL 533  
QY 46 KT-TNMKRLGVDPPCGVLDPKKEAVLLA-----VSCDAFAFGQE 82  
Db 534 ATHQSYQKLDYTPPSNPLPDKFKPILGSNNKPKICLDAYGYGHD 576

RESULT 14  
US-09-074-658-19  
; Sequence 19, Application US/09074658  
; Patent No. 6184371  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M  
; APPLICANT: Run-Pan Du  
; APPLICANT: Quijun Wang  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/074,658  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-795  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1076 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-074-658-19

Query Match 9.1%; Score 62; DB 4; Length 1076;  
Best Local Similarity 22.3%; Pred. No. 63;  
Matches 23; Conservative 15; Mismatches 37; Indels 28; Gaps 4;  
QY 8 DIQTQPGTKIVFNAPYDDKHT-----YHIKV-----INSSARRIGYGI 45  
Db 474 DVNKPFSVKEVDNAYKEQHNLIKAVFNKMKALGSTHHHINLQVGYDKFNSSLSRVEYRL 533  
QY 46 KT-TNMKRLGVDPPCGVLDPKKEAVLLA-----VSCDAFAFGQE 82  
Db 534 ATHQSYQKLDYTPPSNPLPDKFKPILGSNNKPKICLDAYGYGHD 576

RESULT 15  
PCT-US93-11404-2  
; Sequence 2, Application PC/TUS9311404  
; GENERAL INFORMATION:  
; APPLICANT: THE BOARD OF TRUSTEES OF LELAND STANFORD  
; APPLICANT: JUNIOR UNIVERSITY  
; TITLE OF INVENTION: MUCOSAL VASCULAR ADDRESSIN, DNA AND  
; EXPRESSION

; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bertram I. Rowland  
; STREET: 4 Embarcadero Center  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/11404  
; FILING DATE: 23-Nov-1993  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: FP-57452/BIR;STAN-144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 405 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US93-11404-2

Query Match 9.0%; Score 61.5; DB 5; Length 405;  
Best Local Similarity 22.2%; Pred. No. 18;  
Matches 37; Conservative 15; Mismatches 52; Indels 63; Gaps 9;  
QY 2 QSVPPGDIQTQPGTKIVFNA-PYDDKHTYHIKVINSSARRIGYGIKTNM----- 50  
Db 215 KEIPVLQSQTSKPPNTTSAEPY-----ILTSSTAEAVSTGLNITLPSAPPYPKLS 267  
QY 51 -KRLGVDPPCGVLDPK-----EA---VLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101  
Db 268 PRTLSEGPC---RPKIHQDLEAGWELLCEASC-----GPGVTVRWTLAPGDLA 313  
QY 102 RQFRRE-----WFQ-----GDGMVRRKNLPIEYNP 126  
Db 314 TYHKREAGAAWLSVLPFGPMVEGWFCRQDRDQPGQVTNLYVPGQVTP 360

Search completed: September 20, 2002, 09:24:08  
Job time: 86 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 16.56 Seconds  
(without alignments)  
731.114 Million cell updates/sec

Title: US-09-863-063-2  
Perfect score: 682  
Sequence: 1 AQSVPFGDIQTQPGTKIVFN.....EFQGDGMVRRKKNLPLEYNP 126

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682	100.0	127	T16183	hypothetical prote
2	679	99.6	127	G88145	protein F58A6.8 [i
3	679	99.6	127	A88165	protein ZK1248.6 [
4	679	99.6	127	G88686	protein msp-19 [im
5	679	99.6	127	C88688	protein msp-113 [i
6	679	99.6	127	H88688	protein msp-59 [im
7	679	99.6	127	B88689	protein msp-65 [im
8	679	99.6	127	C88689	protein msp-51 [im
9	679	99.6	127	H88792	protein K07F5.1 [i
10	679	99.6	127	H88146	protein C34F11.4 [
11	679	99.6	127	E88134	protein msp-40 [im
12	679	99.6	127	F88138	protein MSP-31 [im
13	679	99.6	127	D88164	protein msp-142 [i
14	678	99.4	127	T24885	hypothetical prote
15	675	99.0	127	T21640	hypothetical prote
16	675	99.0	127	T16684	major sperm protei
17	675	99.0	127	A88139	protein ZK546.6 [i
18	675	99.0	133	T27902	hypothetical prote
19	673	98.7	127	F88146	protein C34F11.6 [
20	671	98.4	127	C88164	protein K05F1.7 [i
21	670	98.2	127	A88683	protein C09B9.6 [i
22	668	97.9	127	F88801	protein C04G2.4 [i
23	665	97.5	127	T23486	hypothetical prote
24	639	93.7	190	T16687	major sperm protei
25	590	86.5	127	B45528	major sperm protei
26	585	85.8	127	A45528	major sperm protei
27	574	84.2	127	A45944	major sperm protei
28	428	62.8	484	T26393	hypothetical prote
29	418	61.3	77	F88165	protein ZK1248.4 [

30	354.5	52.0	95	2	T34500	hypothetical prote
31	351	51.5	169	2	T31484	hypothetical prote
32	302	44.3	99	2	T31606	hypothetical prote
33	218	32.0	418	2	T19800	hypothetical prote
34	134	19.6	260	2	A57245	VAMP-binding prote
35	116.5	17.1	37	2	T29808	hypothetical prote
36	95	13.9	243	2	JG0186	vesicle-associated
37	90.5	13.3	245	2	T32774	hypothetical prote
38	84.5	12.4	185	2	T25098	hypothetical prote
39	82.5	12.1	319	2	T27116	hypothetical prote
40	79	11.6	352	2	AB2243	hypothetical prote
41	78.5	11.5	601	2	C81414	penicillin-binding
42	77.5	11.4	4006	2	T09070	probable tenascin
43	77	11.3	612	2	F87690	response regulator
44	76.5	11.2	221	2	T19317	hypothetical prote
45	75.5	11.1	1101	2	T21062	hypothetical prote

ALIGNMENTS

RESULT 1

T16183

hypothetical protein F26G1.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 26-May-2000

C:Accession: T16183

R:Chissoe, S.

submitted to the EMBL Data Library, July 1995/

A:Description: The sequence of C. elegans cosmid F26G1.

A:Reference number: Z18472

A:Accession: T16183

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-127 <CHI>

A:Cross-references: EMBL:U23519; NID:g746524; PID:g746531; PIDN:AAC46807.1; CESP:F26G

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F26G1.7

C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 100.0%; Score 682; DB 2; Length 127;  
Best Local Similarity 100.0%; Pred. No. 8.6e-65;  
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRRIGYGIKTTNMKRLGVDPGCG 60

|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRRIGYGIKTTNMKRLGVDPGCG 61

Qy 61 VLDPKRAVLLAVSCDAFAFGQEDTNDRTITVEWNTPDGAARQFRREWFQGDGMVRRKNL 120

|||||  
Db 62 VLDPKRAVLLAVSCDAFAFGQEDTNDRTITVEWNTPDGAARQFRREWFQGDGMVRRKNL 121

Qy 121 PIEYNP 126

|||||  
Db 122 PIEYNP 127

RESULT 2

G88145

protein F58A6.8 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C:Accession: G88145

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C-

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: G88145

A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-127 <STO>  
A:Cross-references: GB:chr\_II; PIDN:AA96204.1; PID:g1255857; GSPDB:GN00020; CESP:F58A6.8  
C:Genetics:  
A:Gene: F58A6.8  
A:Map position: 2  
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61  
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127

RESULT 3  
A88165  
protein ZK1248.6 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: A88165  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A88165  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-127 <STO>  
A:Cross-references: GB:chr\_II; PID:g862495; GSPDB:GN00020; CESP:ZK1248.6  
A:Note: similar to C. elegans major sperm protein  
C:Genetics:  
A:Gene: ZK1248.6  
A:Map position: 2  
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61  
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127

RESULT 4  
G88686  
protein msp-19 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: G88686

R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_-  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
A:Accession: G88686  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-127 <STO>  
A:Cross-references: GB:chr\_IV; PIDN:AAC26926.1; PID:g3329619; GSPDB:GN00022  
C:Genetics:  
A:Gene: msp-19  
A:Map position: 4  
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61  
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127

RESULT 5  
C88688  
protein msp-113 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: C88688  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_-  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
A:Accession: C88688  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-127 <STO>  
A:Cross-references: GB:chr\_IV; PIDN:AAB42255.1; PID:g1825633; GSPDB:GN00022  
C:Genetics:  
A:Gene: msp-113  
A:Map position: 4  
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61  
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127

## RESULT 6

H88688

protein msp-59 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C;Accession: H88688

R;anonymouse, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: H88688

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-127 &lt;STO&gt;

A;Cross-references: GB:chr\_IV; PIDN:AAB42253.1; PID:g1825631; GSPDB:GN00022

C;Genetics:

A;Gene: msp-59

A;Map position: 4

C;Superfamily: Caenorhabditis elegans major sperm protein

## Query Match

Best Local Similarity 99.6%; Score 679; DB 2; Length 127;

Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK

QY 121 PIEYNP 126

Db 122 PIEYNP 127

## RESULT 7

B88689

protein msp-65 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C;Accession: B88689

R;anonymouse, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: B88689

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-127 &lt;STO&gt;

A;Cross-references: GB:chr\_IV; PIDN:AAB42256.1; PID:g1825634; GSPDB:GN00022

C;Genetics:

A;Gene: msp-65

A;Map position: 4

C;Superfamily: Caenorhabditis elegans major sperm protein

## Query Match

Best Local Similarity 99.6%; Score 679; DB 2; Length 127;

Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK 121

QY 121 PIEYNP 126

Db 122 PIEYNP 127

## RESULT 8

C88689

protein msp-51 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C;Accession: C88689

R;anonymouse, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A;Accession: C88689

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-127 &lt;STO&gt;

A;Cross-references: GB:chr\_IV; PIDN:AAB42254.1; PID:g1825632; GSPDB:GN00022

C;Genetics:

A;Gene: msp-51

A;Map position: 4

C;Superfamily: Caenorhabditis elegans major sperm protein

## Query Match

Best Local Similarity 99.6%; Score 679; DB 2; Length 127;

Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK 120

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRK 121

QY 121 PIEYNP 126

Db 122 PIEYNP 127

## RESULT 9

H88792

protein K07F5.1 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C;Accession: H88792

R;anonymouse, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A;Accession: H88792

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-127 &lt;STO&gt;

A;Cross-references: GB:chr\_IV; PIDN:CAA94282.1; PID:g3878316; GSPDB:GN00022; CESP:K07

C;Genetics:

A;Gene: K07F5.1

A;Map position: 4

C;Superfamily: Caenorhabditis elegans major sperm protein

## Query Match

Best Local Similarity 99.6%; Score 679; DB 2; Length 127;

Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSARRIGYGIKTTNMKRLGVDPCCG 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSARRIGYGIKTTNMKRLGVDPCCG 61

QY 61 VLDPKAEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDGMVRRKNL 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 62 VLDPKAEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDGMVRRKNL 121

QY 121 PIEYNP 126  
| | | | |  
Db 122 PIEYNP 127

RESULT 10  
H88146 protein C34F11.4 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: H88146  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A;Reference number: A75000; PMID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: H88146  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-127 <STO>  
A;Cross-references: GB:chr\_II; PIDN:AA85761.1; PID:g1166627; GSPDB:GN00020; CESP:C34F11.4  
A;Note: similar to major sperm protein  
C:Genetics:  
A;Map position: 2  
A;Gene: C34F11.4  
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSARRIGYGIKTTNMKRLGVDPCCG 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSARRIGYGIKTTNMKRLGVDPCCG 61

QY 61 VLDPKAEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDGMVRRKNL 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 62 VLDPKAEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDGMVRRKNL 121

QY 121 PIEYNP 126  
| | | | |  
Db 122 PIEYNP 127

RESULT 11  
E88134 protein msp-40 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: E88134  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A;Reference number: A75000; PMID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: E88134  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-127 <STO>  
A;Cross-references: GB:chr\_II; PIDN:AA93398.1; PID:g1203940; GSPDB:GN00020; CESP:C33F10.1



A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: D88164  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-127 <STO>  
A;Cross-references: GB:chr\_II; PID:9868174; GSPDB:GN00020; CESP:K05F1.2  
A;Note: K05F1.2  
C;Genetics:  
A;Gene: msp-142  
A;Map position: 2  
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;  
Best Local Similarity 99.2%; Pred. No. 1.8e-64;  
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61  
Qy 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
Qy 121 PIEYNP 126  
Db 122 PIEYNP 127

RESULT 14  
T24885  
hypothetical protein T13F2.11 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 26-May-2000  
C;Accession: T24885  
R;Swinburne, J.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19947  
A;Accession: T24885  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-127 <WIL>  
A;Cross-references: EMBL:Z81122; PIDN:CAB03362.1; GSPDB:GN00022; CESP:T13F2.11  
A;Experimental source: clone T13F2  
C;Genetics:  
A;Gene: CESP:T13F2.11  
A;Map position: 4  
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.4%; Score 678; DB 2; Length 127;  
Best Local Similarity 98.4%; Pred. No. 2.3e-64;  
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61  
Qy 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
Qy 121 PIEYNP 126  
Db 122 PIEYNP 127

RESULT 15  
T21640

hypothetical protein F32B6.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 26-May-2000  
C;Accession: T21640; T24884  
R;Basham, V.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19453  
A;Accession: T21640  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-127 <WIL>  
A;Cross-references: EMBL:Z81074; PIDN:CAB03037.1; GSPDB:GN00022; CESP:F32B6.6  
A;Experimental source: clone F32B6  
R;Swinburne, J.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19947  
A;Accession: T24884  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-127 <WI2>  
A;Cross-references: EMBL:Z81122; PIDN:CAB03361.1; GSPDB:GN00022; CESP:T13F2.10  
A;Experimental source: clone T13F2  
C;Genetics:  
A;Gene: CESP:F32B6.6; CESP:T13F2.10  
A;Map position: 4  
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.0%; Score 675; DB 2; Length 127;  
Best Local Similarity 98.4%; Pred. No. 4.7e-64;  
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61  
Qy 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
Qy 121 PIEYNP 126  
Db 122 PIEYNP 127

Search completed: September 20, 2002, 09:24:30  
Job time: 108 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 11.89 Seconds  
(without alignments)  
410.317 Million cell updates/sec

Title: US-09-863-063-2  
Perfect score: 682  
Sequence: 1 AQSVPFGDIQTQPGTKIVEN.....EWFQDGMVRRKNLPIEYNP 126

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	679	99.6	126	1 MS31_CAEEL	P53017 caenorhabdi
2	675	99.0	126	1 MS33_CAEEL	P53019 caenorhabdi
3	665	97.5	126	1 MS38_CAEEL	P53020 caenorhabdi
4	663	97.2	126	1 MS10_CAEEL	P05634 caenorhabdi
5	658	96.5	126	1 MS56_CAEEL	P05635 caenorhabdi
6	639	93.7	190	1 MS32_CAEEL	P53018 caenorhabdi
7	590	86.5	126	1 MSP2_ASCSU	P27440 ascaris suu
8	590	86.5	126	1 MSP2_ONCVO	P13263 onchocerca
9	585	85.8	126	1 MSP1_ONCVO	P13262 onchocerca
10	574	84.2	126	1 MSP1_ASCSU	P27439 ascaris suu
11	436	63.9	125	1 MSP2_GLORO	P53022 globodera r
12	429	62.9	125	1 MSP1_GLORO	P53021 globodera r
13	425	62.3	125	1 MSP3_GLORO	P53023 globodera r
14	134	19.6	260	1 VP33_APLCA	Q16943 aplysia cal
15	74	10.9	131	1 YOD1_CAEEL	P34593 caenorhabdi
16	73	10.7	179	1 YNC5_CAEEL	P34538 caenorhabdi
17	73	10.7	244	1 SCS2_YEAST	P40075 saccharomyc
18	71	10.4	285	1 YRJ1_STRCO	P41108 streptomyce
19	69	10.1	863	1 MCM4_XENLA	P30564 xenopus lae
20	68.5	10.0	237	1 VT4_SFVKA	P25948 shope fibro
21	68	10.0	221	1 SEGA_BPT4	P32286 bacterioph
22	67.5	9.9	996	1 VGNM_RCMV	P13561 red clover
23	67	9.8	328	1 RPOA_RCMV	P74963 shewanella
24	67	9.8	436	1 SLS6_BRAOL	P07761 brassica ol
25	67	9.8	1630	1 PTP1_DROME	P35992 drosophila
26	67	9.8	4289	1 TENX_HUMAN	P22105 homo sapien
27	66	9.7	616	1 SPAN_STRPU	P98068 strongyloce
28	65	9.5	513	1 PPT1_YEAST	P53043 saccharomyc
29	65	9.5	1031	1 RAD2_YEAST	P07276 saccharomyc
30	64.5	9.5	333	1 LIPB_ARATH	O23021 arabidopsis
31	64	9.4	553	1 PELW_ERWCH	Q05526 erwinia chr
32	64	9.4	2875	1 RRPL_TSWV1	P28976 tomato spot
33	64	9.4	3067	1 CALC_MOUSE	Q60847 mus musculu

34	63.5	9.3	333	1 TAL1_KLUJLA	P34214 kluyveromyc
35	63.5	9.3	1239	1 TOP2_CRIFA	P27570 crithidia f
36	63	9.2	895	1 YA54_METJA	Q58454 methanococc
37	63	9.2	1080	1 MI15_CAEEL	Q23356 caenorhabdi
38	62.5	9.2	337	1 G3P_USTMA	P09317 ustilago ma
39	62.5	9.2	503	1 IMDH_TRIFO	P50097 tritrichomo
40	62	9.1	322	1 YJ96_CAEEL	P49048 caenorhabdi
41	62	9.1	559	1 SUCP_ECOLI	P76041 escherichia
42	61.5	9.0	447	1 SUN_BACSU	P94464 bacillus su
43	61.5	9.0	818	1 PLD2_ORYSA	P93844 oryza sativ
44	61	8.9	234	1 ENTD_SALTY	Q56064 salmonella
45	61	8.9	378	1 FENS_ORYSA	P41345 oryza sativ

ALIGNMENTS

RESULT 1					
MS31_CAEEL	MS31_CAEEL	STANDARD;	PRT;	126 AA.	
ID	MS31_CAEEL	STANDARD;	PRT;	126 AA.	
AC	P53017;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Major sperm protein 31/40/142 (MSP).				
GN	(MSP-31 OR R05F9.13) AND (MSP-40 OR C33F10.9) AND (MSP-51 OR ZK354.5)				
GN	AND (MSP-59 OR ZK354.11) AND (MSP-65 OR ZK354.1) AND (MSP-113 OR				
GN	ZK354.4) AND (MSP-142 OR K05F1.2) AND C34F11.4 AND F58A6.8 AND K07F5.1				
GN	AND ZK1248.6.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OC	Rhabditidae; Peloderinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Hallsworth K.;				
RL	Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Waterston R.;				
RL	Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Wohldmann P.;				
RL	Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Bentley D.;				
RL	Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Waterston R.;				
RL	Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.				
RN	[6]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Hembry C.;				
RL	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.				
RN	[7]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2;				
RA	Latreille P.;				
RL	Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.				
RN	[8]				
RP	SEQUENCE FROM N.A. (MSP-51; MSP-59; MSP-65 AND MSP-113).				
RC	STRAIN=BRISTOL N2;				
RA	Johnson D., Wamsley P., Bradshaw H.;				
RL	Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.				





```
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68879; CA93089.1; -.
CC HSSP; P27439; 3MSP.
CC WormPep; K08F4.8; CE06156.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 126 AA; 14064 MW; 98F888DFD8BB7A49 CRC64;
CC -----
CC Query Match 97.5%; Score 665; DB 1; Length 126;
CC Best Local Similarity 97.6%; Pred. No. 1.5e-64;
CC Matches 123; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
DB 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
DB 61 VFDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
QY 121 PIEYNP 126
DB 121 PIEYNP 126
RESULT 4
MS10_CAEEL STANDARD; PRT; 126 AA.
AC P05634;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 10 (MSP).
GN MSP-10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-84191131; PubMed-6325882;
RA Klass M.R., Kinsley S., Lopez L.C.;
RT "Isolation and characterization of a sperm-specific gene family in
RL the nematode Caenorhabditis elegans.";
RL Mol. Cell. Biol. 4:529-537(1984).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68879; CA93089.1; -.
CC HSSP; P27439; 3MSP.
CC WormPep; K08F4.8; CE06156.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 126 AA; 14064 MW; 98F888DFD8BB7A49 CRC64;
CC -----
CC Query Match 97.5%; Score 665; DB 1; Length 126;
CC Best Local Similarity 97.6%; Pred. No. 1.5e-64;
CC Matches 123; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
DB 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
DB 61 VFDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
QY 121 PIEYNP 126
DB 121 PIEYNP 126
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02617; AAA28115.1; -.
CC HSSP; P27439; 3MSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 126 AA; 14093 MW; 835CA9DBFD80D5D9 CRC64;
CC -----
CC Query Match 97.2%; Score 663; DB 1; Length 126;
CC Best Local Similarity 96.8%; Pred. No. 2.5e-64;
CC Matches 122; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC -----
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
DB 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
DB 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120
QY 121 PIEYNP 126
DB 121 PIEYNP 126
RESULT 5
MS56_CAEEL STANDARD; PRT; 126 AA.
AC P05635;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 56 (MSP).
GN MSP-56.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-84191131; PubMed-6325882;
RA Klass M.R., Kinsley S., Lopez L.C.;
RT "Isolation and characterization of a sperm-specific gene family in
RL the nematode Caenorhabditis elegans.";
RL Mol. Cell. Biol. 4:529-537(1984).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02618; AAA28116.1; -.
CC HSSP; P27439; 3MSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 126 AA; 14063 MW; 82F7B36A4D80C5C8 CRC64;
CC -----
```



```
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89365002; PubMed=2770787;
RA Scott A.L., Dinman J., Sussman D.J., Yenbutr P., Ward S.;
RT "Major sperm protein genes from Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 36:119-126(1989).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04663; AAA29421.1; -.
DR PIR; B45528; B45528.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT SEQUENCE 126 AA; 14225 MW; D40F9552B561E9D4 CRC64;
SQ
Query Match 86.5%; Score 590; DB 1; Length 126;
Best Local Similarity 84.8%; Pred. No. 1.7e-56;
Matches 106; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AQSVPGGDIHTQPGSKIVFNAPYDDKHTYHIKVINAGRRIGWAIKTTNMKRLGVDPPCG 60

QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 VLDPKENVLMVSCDTFDATREDINNDRTITIEWTNTPDGAARQFRREWFQDGMVRRKNL 120

QY 121 PIEYN 125
   |||||
Db 121 PIEYN 125

RESULT 9
MSPL_ONCVO
ID MSPL_ONCVO STANDARD; PRT; 126 AA.
AC P13262;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Major sperm protein 1 (MSPl).
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89365002; PubMed=2770787;
RA Scott A.L., Dinman J., Sussman D.J., Yenbutr P., Ward S.;
RT "Major sperm protein genes from Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 36:119-126(1989).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04663; AAA29421.1; -.
DR PIR; B45528; B45528.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT SEQUENCE 126 AA; 14225 MW; D40F9552B561E9D4 CRC64;
SQ
Query Match 85.8%; Score 585; DB 1; Length 126;
Best Local Similarity 84.0%; Pred. No. 5.9e-56;
Matches 105; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AQSVPGGDIHTQPGSKIVFNAPYDDKHTYHIKVINAGRRIGWAIKTTNMKRLGVDPPCG 60

QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 VLDPKENVLMVSCDTFDATREDINNDRTITIEWTNTPDGAARQFRREWFQDGMVRRKNL 120

QY 121 PIEYN 125
   |||||
Db 121 PIEYN 125

RESULT 10
MSPL_ASCSU
ID MSPL_ASCSU STANDARD; PRT; 126 AA.
AC P27439; P27441;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major sperm protein, isoform alpha (Alpha-MSp).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RC TISSUE=Sperm;
RX MEDLINE=92407055; PubMed=1527183;
RA King K.L., Stewart M., Roberts T.M., Seavy M.;
RT "Structure and macromolecular assembly of two isoforms of the major
RT sperm protein (MSP) from the amoeboid sperm of the nematode, Ascaris
RT suum.";
RL J. Cell Sci. 101:847-857(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87031211; PubMed=3770294;
RA Bennett K.L., Ward S.;
RT "Neither a germ line-specific nor several somatically expressed genes
RT are lost or rearranged during embryonic chromatin diminution in the
RT nematode Ascaris lumbricoidea var. suum.";
RL Dev. Biol. 118:141-147(1986).
RN [3]
RP SEQUENCE FROM N.A.
RA Bullock T.L., Parathasathy G., King K.L., Kent M.L., Roberts T.M.,
RA Stewart M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97070381; PubMed=8913307;
```



RA Bullock T.L., Roberts T.M., Stewart M.;  
RT "2.5-A resolution crystal structure of the motile major sperm protein  
RL (MSP) of Ascaris suum.";  
RN J. Mol. Biol. 263:284-296(1996).  
RP [5]  
RX STRUCTURE BY NMR.  
RA MEDLINE=99096891; PubMed=9878374;  
RA Haaf A., Leclaire L. III, Roberts G., Kent H.M., Roberts T.M.,  
RA Stewart M., Neuhaus D.;  
RT "Solution structure of the motile major sperm protein (MSP) of  
RT Ascaris suum - evidence for two manganese binding sites and the  
RT possible role of divalent cations in filament formation.";  
RL J. Mol. Biol. 284:1611-1624(1998).  
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING  
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS  
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.  
CC -!- SUBUNIT: FORMS FILAMENTS 10 NM WIDE, WITH A CHARACTERISTIC  
CC SUBSTRUCTURE REPEATING AXIALLY AT 9 NM.  
CC -!- TISSUE SPECIFICITY: SPERM.  
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M15680; AAA29375.1; -.  
DR EMBL; X94249; CAA63933.1; -.  
DR PIR; A45944; A45944.  
DR PDB; 1MSP; 07-DEC-96.  
DR PDB; 3MSP; 20-APR-99.  
DR InterPro; IPR000535; MSP\_domain.  
DR Pfam; PF00635; MSP\_domain; 1.  
KW Cytoskeleton; Acetylation; Sperm; Multigene family.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT CONFLICT 113 113 G -> D (IN REF. 1).  
SQ SEQUENCE 126 AA; 14259 MW; 477DCEF6F4CFDD8F CRC64;  
  
Query Match 84.2%; Score 574; DB 1; Length 126;  
Best Local Similarity 82.4%; Pred. No. 8.9e-55;  
Matches 103; Conservative 7; Mismatches 15; Indels 0; Gaps 0;  
  
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60  
Db 1 AQSVPFGDINTQPSQKIVFNAPYDDKHHTYHIKVINAGRRIGWAIKTTNMRRLSVDPDPCG 60  
  
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
Db 61 VLDPKKVLMAVSCDTFNATEDLNDRITIEWTNTPDGAAKQFRREWFQDGMVRRKNL 120  
  
QY 121 PIEYN 125  
Db 121 PIEYN 125  
  
RESULT 11  
MSP2\_GLORO  
ID MSP2\_GLORO STANDARD; PRT; 125 AA.  
AC P53022;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Major sperm protein 2.  
GN MSP-2.  
OS Globodera rostochiensis (Golden nematode).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
OX NCBI\_TaxID=31243;

RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RO1;  
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,  
RA McPherson M.J.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING  
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS  
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY  
CC SIMILARITY).  
CC -!- TISSUE SPECIFICITY: SPERM.  
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L24500; AAA29147.1; -.  
DR HSP; P27439; 1MSP.  
DR InterPro; IPR000535; MSP\_domain.  
DR Pfam; PF00635; MSP\_domain; 1.  
KW Cytoskeleton; Acetylation; Sperm; Multigene family.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
SQ SEQUENCE 125 AA; 13801 MW; 1B04AD5756511B13 CRC64;  
  
Query Match 63.9%; Score 436; DB 1; Length 125;  
Best Local Similarity 63.9%; Pred. No. 6e-40;  
Matches 78; Conservative 16; Mismatches 28; Indels 0; Gaps 0;  
  
QY 4 VPPGDIQTQPGTKIVFNAPYDDKHHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVLD 63  
Db 3 LPPGDIATMPNQKVVFNAPFDNKAITYYRVINPGTNRIGFAFTKPKRINMPPNGVLG 62  
  
QY 64 PKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNLPIE 123  
Db 63 PKESVNVAVISCDADFDPSSSEDTKGDRTVVEWCNTPDPAATAAFKLEWFQDGMVRRKNLPIE 122  
  
QY 124 YN 125  
Db 123 YN 124  
  
RESULT 12  
MSP1\_GLORO  
ID MSP1\_GLORO STANDARD; PRT; 125 AA.  
AC P53021;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Major sperm protein 1.  
GN MSP-1.  
OS Globodera rostochiensis (Golden nematode).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
OX NCBI\_TaxID=31243;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RO1;  
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,  
RA McPherson M.J.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING  
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS  
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY  
CC SIMILARITY).  
CC -!- TISSUE SPECIFICITY: SPERM.  
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.





Search completed: September 20, 2002, 09:23:35  
Job time: 53 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 24.67 Seconds  
(without alignments)  
883.557 Million cell updates/sec

Title: US-09-863-063-2  
Perfect score: 682  
Sequence: 1 AQSVPGGDIQTQPGTKIVN.....ENFQDGMVRRKNLPIEYNP 126

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- SPTREMBL\_19:\*\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	582	100.0	127	5	Q19832	Q19832 caenorhabdi
2	679	99.6	130	5	Q9N2M1	Q9n2m1 caenorhabdi
3	678	99.4	127	5	Q94053	Q94053 caenorhabdi
4	676	99.1	127	5	Q95XU7	Q95xu7 caenorhabdi
5	675	99.0	127	5	Q9TVW5	Q9tvw5 caenorhabdi
6	675	99.0	127	5	Q23519	Q23519 caenorhabdi
7	675	99.0	133	5	Q23524	Q23524 caenorhabdi
8	673	98.7	127	5	Q18461	Q18461 caenorhabdi
9	671	98.4	127	5	Q21244	Q21244 caenorhabdi
10	670	98.2	127	5	Q17856	Q17856 caenorhabdi
11	668	97.9	127	5	Q27280	Q27280 caenorhabdi
12	498	73.0	107	5	Q9GNV8	Q9gnv8 onchocerca
13	496	72.7	107	5	Q9GNV9	Q9gnv9 onchocerca
14	494	72.4	107	5	Q9GNW0	Q9gnw0 onchocerca
15	493	72.3	107	5	Q9GN10	Q9gni0 mansonella
16	492	72.1	107	5	Q9GNW2	Q9gnw2 onchocerca

17	490	71.8	107	5	Q9GNW6	Q9gnw6 mansonella
18	488	71.6	107	5	Q9GNX0	Q9gnx0 mansonella
19	488	71.6	107	5	Q9GNW7	Q9gnw7 mansonella
20	487	71.4	107	5	Q9GNX1	Q9gnx1 mansonella
21	485	71.1	107	5	Q9GNW1	Q9gnw1 onchocerca
22	485	71.1	107	5	Q9GNW9	Q9gnw9 mansonella
23	484	71.0	107	5	Q9GNW8	Q9gnw8 mansonella
24	484	71.0	107	5	Q9GNW4	Q9gnw4 mansonella
25	483	70.8	107	5	Q9GNW5	Q9gnw5 mansonella
26	428	62.8	442	5	Q9NAP2	Q9nap2 caenorhabdi
27	428	62.8	484	5	Q9NAM2	Q9nam2 caenorhabdi
28	418	61.3	77	5	Q95PJ7	Q95pj7 caenorhabdi
29	357	52.3	90	5	Q26098	Q26098 pratylenchu
30	354.5	52.0	95	5	Q23428	Q23428 caenorhabdi
31	351	51.5	167	5	Q9U2V8	Q9u2v8 caenorhabdi
32	351	51.5	169	5	Q9U2W1	Q9u2w1 caenorhabdi
33	339	49.7	88	5	Q27405	Q27405 pratylenchu
34	329	48.2	85	5	Q26097	Q26097 pratylenchu
35	309	45.3	84	5	Q26111	Q26111 pratylenchu
36	305	44.7	83	5	Q26096	Q26096 pratylenchu
37	302	44.3	99	5	Q9NAF5	Q9naf5 caenorhabdi
38	293	43.0	83	5	Q26112	Q26112 pratylenchu
39	279	40.9	141	5	Q26316	Q26316 dictyocaulu
40	218	32.0	418	5	Q18503	Q18503 caenorhabdi
41	116.5	17.1	37	5	Q18145	Q18145 caenorhabdi
42	114	16.7	242	4	Q9UBZ2	Q9ubz2 homo sapien
43	114	16.7	242	4	Q75453	Q75453 homo sapien
44	114	16.7	242	4	Q9P0L0	Q9p0l0 homo sapien
45	114	16.7	242	11	Q9Z270	Q9z270 rattus norv

ALIGNMENTS

RESULT	1
Q19832	
ID	Q19832 PRELIMINARY; PRT; 127 AA.
AC	Q19832;
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE	HYPOTHETICAL 14.2 KDA PROTEIN.
GN	F26G1.7.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RX	MEDLINE=99069613; PubMed=9851916;
RA	None;
RT	"Genome sequence of the nematode C. elegans: a platform for
RT	investigating biology. The C. elegans Sequencing Consortium.";
RL	Science 282:2012-2018(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RA	Chissee S.;
RT	"The sequence of C. elegans cosmid F26G1.";
RL	Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RA	Waterston R.;
RT	"Direct Submission.";
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; U23519; AAK31504.1; -.
DR	HSSP; P27439; 3MSP.
DR	InterPro; IPR000535; MSP_domain.
DR	Pfam; PF00635; MSP_domain; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 127 AA; 14237 MW; 71671F31BEA5B147 CRC64;





RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024841; AAK68539.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 127 AA; 14195 MW; 69EA78C3A24B4347 CRC64;

Query Match 99.1%; Score 676; DB 5; Length 127;  
Best Local Similarity 98.4%; Pred. No. 3.2e-65;  
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRVGVDPGCG 61  
|||||

QY 61 VLPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 121  
|||||

QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127  
|||||

RESULT 5  
Q9TVW5 PRELIMINARY; PRT; 127 AA.  
AC Q9TVW5;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE T13F2.10 PROTEIN.  
GN T13F2.10 OR F32B6.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Swinburne J.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Basham V.M.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z81122; CAB03361.1; -.  
DR EMBL; Z81074; CAB03037.1; -.  
DR HSSP; P27439; 3MSP.  
DR InterPro; IPR000535; MSP\_domain.  
DR Pfam; PF00635; MSP\_domain; 1.  
SQ SEQUENCE 127 AA; 14181 MW; 69F6962335A5B147 CRC64;

Query Match 99.0%; Score 675; DB 5; Length 127;  
Best Local Similarity 98.4%; Pred. No. 4.1e-65;  
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60  
|||||  
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61  
|||||

QY 61 VLPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||

Db 62 VLPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMARRKNL 121  
QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127  
|||||

RESULT 6  
Q23519 PRELIMINARY; PRT; 127 AA.  
ID Q23519  
AC Q23519;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 14.2 KDA PROTEIN.  
GN ZK546.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Hallsworth K.;  
RT "The sequence of C. elegans cosmid ZK546.";  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U29380; AAA68739.1; -.  
DR HSSP; P27439; 3MSP.  
DR InterPro; IPR000535; MSP\_domain.  
DR Pfam; PF00635; MSP\_domain; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 127 AA; 14239 MW; 69F6962DCEACC152 CRC64;

Query Match 99.0%; Score 675; DB 5; Length 127;  
Best Local Similarity 99.2%; Pred. No. 4.1e-65;  
Matches 124; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCGV 61  
|||||  
Db 3 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCGV 62  
|||||

QY 62 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNLP 121  
|||||  
Db 63 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNLP 122  
|||||

QY 122 IEYNP 126  
|||||  
Db 123 IEYNP 127  
|||||

RESULT 7  
Q23524 PRELIMINARY; PRT; 133 AA.  
ID Q23524  
AC Q23524;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 14.9 KDA PROTEIN.



SQ SEQUENCE 127 AA; 14228 MW; 7D5D628EF5511AA7 CRC64;

Query Match 98.4%; Score 671; DB 5; Length 127;  
Best Local Similarity 98.4%; Pred. No. 1.1e-64;  
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 60  
|||||  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYRIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 61  
|||||

QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 121  
|||||

QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127  
|||||

RESULT 10  
Q17856  
ID Q17856 PRELIMINARY; PRT; 127 AA.  
AC Q17856;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE C09B9.6 PROTEIN (HYPOTHETICAL PROTEIN R13H9.2).  
GN C09B9.6 OR R13H9.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Bradshaw H.;  
RT "The sequence of C. elegans cosmid C09B9.";  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA R, Wilson, Bradshaw H.;

RT "The sequence of C. elegans cosmid R13H9.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U50069; AAB37555.1; -.  
DR EMBL; AC006681; AAK85492.1; -.  
DR HSSP; P27439; 3MSP.  
DR InterPro; IPR000535; MSP\_domain.  
DR Pfam; PF00635; MSP\_domain; 1.  
SQ SEQUENCE 127 AA; 14251 MW; 28E69731A2D5AD2B CRC64;

Query Match 98.2%; Score 670; DB 5; Length 127;  
Best Local Similarity 98.4%; Pred. No. 1.4e-64;  
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 60  
|||||  
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 61  
|||||

QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 120  
|||||  
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKNL 121  
|||||

QY 121 PIEYNP 126  
|||||  
Db 122 PIEYNP 127  
|||||

RESULT 11  
Q27280  
ID Q27280 PRELIMINARY; PRT; 127 AA.  
AC Q27280;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE C04G2.4 PROTEIN.  
GN C04G2.4 OR ZK1251.6 OR K07F5.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hembry C.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA McMurray A.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Hembry C.;

RL	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.										
DR	EMBL; Z70718; CAA94674.1; -										
DR	EMBL; Z68222; CAA92502.1; -										
DR	EMBL; Z70284; CAA94278.1; -										
DR	EMBL; Z70284; CAA94283.1; -										
DR	HSSP; P27439; 3MSP.										
DR	InterPro; IPR000535; MSP_domain.										
DR	Pfam; PF00635; MSP_domain; 1.										
SQ	SEQUENCE 127 AA; 14236 MW; E5B96631BEBF1419 CRC64;										
Query Match 97.9%; Score 668; DB 5; Length 127;											
Best Local Similarity 97.6%; Pred. No. 2.3e-64;											
Matches 123; Conservative 1; Mismatches 2; Indels 0; Gaps 0;											
QY	1	AQSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG	60								
Db	2	AQSVPPGDIQTQPNKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG	61								
QY	61	VLPDKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL	120								
Db	62	VLPDKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQDGMVRRKNL	121								
QY	121	PIEYNP 126									
Db	122	PIEYNP 127									
RESULT	12										
Q9GNV8											
ID	Q9GNV8	PRELIMINARY;	PRT;	107	AA.						
AC	Q9GNV8;										
DT	01-MAR-2001 (TremBLrel. 16, Created)										
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)										
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)										
DE	MAJOR SPERM PROTEIN (FRAGMENT).										
GN	MSP.										
OS	Onchocerca volvulus.										
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;										
OC	Onchocercidae; Onchocerca.										
OX	NCBI_TaxID=6282;										
RN	[1]										
RP	SEQUENCE FROM N.A.										
RC	STRAIN=OVNODBRA2C5 ;										
RX	MEDLINE=21321180; PubMed=11428336;										
RA	Morales Hojas R., Post R.J.;										
RT	"Regional genetic variation in the major sperm protein genes of										
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";										
RL	Int. J. Parasitol. 30:1459-1465(2000).										
DR	EMBL; AJ404208; CAC20742.1; -										
DR	HSSP; P27439; 1MSP.										
DR	InterPro; IPR000535; MSP_domain.										
DR	Pfam; PF00635; MSP_domain; 1.										
FT	NON_TER 1										
FT	NON_TER 107										
SQ	SEQUENCE 107 AA; 11987 MW; 0C4325F7F6E5DF07 CRC64;										
Query Match 73.0%; Score 498; DB 5; Length 107;											
Best Local Similarity 82.2%; Pred. No. 4e-46;											
Matches 88; Conservative 8; Mismatches 11; Indels 0; Gaps 0;											
QY	7	GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKE	66								
Db	1	GDIHTQPGSKIVFNAPYDDKHTYHIKITNAGRRIGWAIKTTNMKRLGVDPPCGVLDPKE	60								
QY	67	AVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDG	113								
Db	61	NVLMAVSCDTFDAAREIDNDRITVEWTNTPDGAARQFRREWFQGDG	107								
RESULT	13										
Q9GNV9											
ID	Q9GNV9	PRELIMINARY;	PRT;	107	AA.						
AC	Q9GNV9;										
DT	01-MAR-2001 (TremBLrel. 16, Created)										
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)										
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)										
DE	MAJOR SPERM PROTEIN (FRAGMENT).										
GN	MSP.										
OS	Onchocerca volvulus.										
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;										
OC	Onchocercidae; Onchocerca.										
OX	NCBI_TaxID=6282;										
RN	[1]										
RP	SEQUENCE FROM N.A.										
RC	STRAIN=OVNODBRA2C5 ;										
RX	MEDLINE=21321180; PubMed=11428336;										
RA	Morales Hojas R., Post R.J.;										
RT	"Regional genetic variation in the major sperm protein genes of										
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";										
RL	Int. J. Parasitol. 30:1459-1465(2000).										
DR	EMBL; AJ404208; CAC20742.1; -										
DR	HSSP; P27439; 1MSP.										
DR	InterPro; IPR000535; MSP_domain.										
DR	Pfam; PF00635; MSP_domain; 1.										
FT	NON_TER 1										
FT	NON_TER 107										
SQ	SEQUENCE 107 AA; 11987 MW; 0C4325F7F6E5DF07 CRC64;										

ID	Q9GNV9	PRELIMINARY;	PRT;	107	AA.
AC	Q9GNV9;				
DT	01-MAR-2001 (TremBLrel. 16, Created)				
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)				
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)				
DE	MAJOR SPERM PROTEIN (FRAGMENT).				
GN	MSP.				
OS	Onchocerca volvulus.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;				
OC	Onchocercidae; Onchocerca.				
OX	NCBI_TaxID=6282;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=OVNODBRA2C4 ;				
RX	MEDLINE=21321180; PubMed=11428336;				
RA	Morales Hojas R., Post R.J.;				
RT	"Regional genetic variation in the major sperm protein genes of				
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea)."				
RL	Int. J. Parasitol. 30:1459-1465(2000).				
DR	EMBL; AJ404207; CAC20741.1; -				
DR	HSSP; P27439; 1MSP.				
DR	InterPro; IPR000535; MSP_domain.				
DR	Pfam; PF00635; MSP_domain; 1.				
FT	NON_TER 1				
FT	NON_TER 107				
SQ	SEQUENCE 107 AA; 12017 MW; 0C5E932C86E5DF07 CRC64;				

Query Match 72.7%; Score 496; DB 5; Length 107;										
Best Local Similarity 82.2%; Pred. No. 6.5e-46;										
Matches 88; Conservative 8; Mismatches 11; Indels 0; Gaps 0;										
QY	7	GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPKE	66							
Db	1	GDIHTQPGSKIVFNAPYDDKHTYHIKTNAGRRIGWAIKTTNMKRLGVDPPCGVLDPKE	60							
QY	67	AVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDG	113							
Db	61	NVLMAVSCDTFDATREDINNDRITIEWTNTPDGAARQFRREWFQGDG	107							

RESULT	14														
Q9GNW0															
ID	Q9GNW0	PRELIMINARY;	PRT;	107	AA.										
AC	Q9GNW0;														
DT	01-MAR-2001 (TremBLrel. 16, Created)														
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)														
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)														
DE	MAJOR SPERM PROTEIN (FRAGMENT).														
GN	MSP.														
OS	Onchocerca volvulus.														
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;														
OC	Onchocercidae; Onchocerca.														
OX	NCBI_TaxID=6282;														
RN	[1]														
RP	SEQUENCE FROM N.A.														
RC	STRAIN=OVNODBRA2C3 ;														
RX	MEDLINE=21321180; PubMed=11428336;														
RA	Morales Hojas R., Post R.J.;														
RT	"Regional genetic variation in the major sperm protein genes of														
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";														
RRL	Int. J. Parasitol. 30:1459-1465(2000).														
DR	EMBL; AJ404206; CAC20740.1; -.														
DR	HSSP; P27439; IMSP.														
DR	InterPro; IPR000535; MSP_domain.														
DR	Pfam; PF00635; MSP_domain; 1.														
FT	NON_TER	1													
FT	NON_TER	107	107												
SEQ	SEQUENCE	107	AA;	12018	MW;	OC5E932130E5DF07	CRC64;								



Matches	88;	Conservative	7;	Mismatches	12;	Indels	0;	Gaps	0;
QY	7	GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPK	66						
Db	1	GDIHTQPGSKIVFNAPYDDKHTYHIKITNAGRRIGWAIKTTNMKRLGVDPPCGVLDPK	60						
QY	67	AVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDG	113						
Db	61	NVLMVAVSCDTFDATREDINNDRITIEWTNTPDGAAEQFRREWFQGDG	107						
RESULT 15									
Q9GNI0									
ID	Q9GNI0	PRELIMINARY;	PRT;	107	AA.				
AC	Q9GNI0;								
DT	01-MAR-2001	(TReMBLrel. 16, Created)							
DT	01-MAR-2001	(TReMBLrel. 16, Last sequence update)							
DT	01-DEC-2001	(TReMBLrel. 19, Last annotation update)							
DE	MAJOR SPERM PROTEIN (FRAGMENT).								
GN	MSP.								
OS	Mansonella ozzardi.								
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;								
OC	Onchocercidae; Mansonella.								
OX	NCBI_TaxID=122354;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=VARIOUS STRAINS;								
RX	MEDLINE=21321180; PubMed=11428336;								
RA	Morales Hojas R., Post R.J.;								
RT	"Regional genetic variation in the major sperm protein genes of								
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea)."								
RL	Int. J. Parasitol. 30:1459-1465(2000).								
DR	EMBL; AJ404225;	CAC20724.1;	-						
DR	EMBL; AJ404210;	CAC20709.1;	-						
DR	EMBL; AJ404211;	CAC20710.1;	-						
DR	EMBL; AJ404212;	CAC20711.1;	-						
DR	EMBL; AJ404213;	CAC20712.1;	-						
DR	EMBL; AJ404215;	CAC20714.1;	-						
DR	EMBL; AJ404219;	CAC20718.1;	-						
DR	EMBL; AJ404221;	CAC20720.1;	-						
DR	EMBL; AJ404222;	CAC20721.1;	-						
DR	HSSP; P27439;	1MSP.							
DR	InterPro; IPR000535;	MSP_domain.							
DR	Pfam; PF00635;	MSP_domain; 1.							
FT	NON_TER	1							
FT	NON_TER	107	107						
SQ	SEQUENCE	107	AA;	12017	MW;	1A0BD02C86E5C61E	CRC64;		

Query Match	72.3%;	Score	493;	DB	5;	Length	107;
Best Local Similarity	81.3%;	Pred. No.	1.4e-45;				
Matches	87;	Conservative	9;	Mismatches	11;	Indels	0;
Gaps	0;						
QY	7	GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGVLDPK	66				
Db	1	GDIHTQPGSKIVFNAPYDDKHTYHIKITNAGRRIGWAIKTTNMRLGVDPPCGVLDPK	60				
QY	67	AVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAARQFRREWFQGDG	113				
Db	61	NVLMVAVSCDTFDATKEDINNDRITIEWTNTPDGAAKQFRREWFQGDG	107				

